

Amendments to the Claims:

This listing of the claims will replace all prior versions and listings of the claims in the application:

Listing of Claims:

1. (Currently amended) A noise reduction system including an audio-visual user interface therein for combining visual features extracted from a digital video sequence with audio features extracted from an analog audio sequence including background noise in an environment of a speaker, said noise reduction system comprising:

an audio sequence detection device for detecting said analog audio sequence;

an audio feature extraction and analysis device for analyzing said analog audio sequence and extracting said audio features therefrom;

a video sequence detection device for detecting said video sequence;

a visual feature extraction and analysis device for analyzing the detected video sequence and extracting said visual features therefrom;

a noise reduction circuit configured to separate a speaker's voice from said background noise based on a combination of derived speech characteristics by removing said separated background noise from said analog audio sequence and configured to output a speech activity indication signal comprising a combination of speech activity estimates supplied by said audio feature extraction and analysis device and said visual feature extraction and analysis device; and

a multi-channel acoustic echo cancellation unit configured to perform a near-end speaker detection and double-talk detection algorithm based on the speech characteristics derived by said audio feature extraction and analysis device and said visual feature extraction and analysis device,

wherein said noise reduction circuit is further configured to subtract a discretized version of an estimated noise power density spectrum of said background noise from a discrete signal spectrum of an analog-to-digital converted version of said

analog audio sequence, said estimated noise power density spectrum being based on said audio features, said visual features and said discrete signal spectrum.

2. (Previously presented) A noise reduction system according to claim 1, further comprising:

a device for switching off an audio channel if said speech activity indication signal falls below a predefined threshold value.

3. (Previously presented) A noise reduction system according to claim 1, wherein said audio feature extraction and analysis device comprises an amplitude detector.

4. (Currently amended) A near-end speaker detection method for reducing noise in a detected analog audio sequence, said method comprising:

converting said analog audio sequence into a digital audio sequence;

calculating a corresponding discrete signal spectrum of the digital audio sequence by performing a Fast Fourier Transform (FFT);

detecting a voice of a speaker from said discrete signal spectrum by analyzing visual features extracted from a video sequence associated with extracted and analyzed audio features of the audio sequence, the visual features including current locations of face, lip movements and/or facial expressions of the speaker in a sequence of images in the video sequence;

estimating a noise power density spectrum of statistically distributed background noise based on said audio features, said visual features and said discrete signal spectrum~~a signal that represents the voice of the speaker;~~

subtracting a discretized version of the estimated noise power density spectrum from the discrete signal spectrum of the digital audio sequence to obtain a difference signal; and

calculating a corresponding discrete time-domain signal of the obtained difference signal by performing an Inverse Fast Fourier Transform (IFFT) to provide a recognized speech signal.

5. (Previously presented) A near-end speaker detection method according to claim 4, further comprising:

performing a multi-channel acoustic echo cancellation algorithm which models echo path impulse responses by means of adaptive finite impulse response (FIR) filters and subtracts echo signals from the analog audio sequence based on acoustic-phonetic speech characteristics derived by an algorithm for extracting the visual features from the video sequence associated with the audio sequence and including the locations of the face, lip movements and/or facial expressions of the speaker in a sequence of images in the video sequence.

6. (Previously presented) A near-end speaker detection method according to claim 5, wherein said multi-channel acoustic echo cancellation algorithm performs a double-talk detection procedure.

7. (Previously presented) A near-end speaker detection method according to claim 4, wherein said acoustic-phonetic speech characteristics are based on detecting opening of a mouth of the speaker as an estimate of acoustic energy of articulated vowels and/or diphthongs, detecting rapid movement of the lips of the speaker as a hint to labial or labio-dental consonants, and/or detecting other phonetic characteristics associated with position and movement of the lips and/or voice and/or pronunciation of said speaker.

8. (Previously presented) A near-end speaker detection method according to claim 4, wherein detecting the voice of said speaker comprises:

detecting the voice of said speaker from the discrete signal spectrum of the digital audio sequence using a learning procedure by analyzing the visual features extracted from the video sequence associated with the audio sequence and including the current locations of the face, lip movements and/or facial expressions of the speaker in a sequence of images in the video sequence.

9. (Previously presented) A near-end speaker detection method according to claim 4, further comprising:

correlating the discrete signal spectrum of a delayed version of the digital audio signal with an audio speech activity estimate obtained by amplitude detection of a band-pass-filtered discrete signal spectrum to provide an estimate for a frequency spectrum corresponding to a signal which represents a voice of said speaker as well as an estimate for the noise power density spectrum of the statistically distributed background noise.

10. (Previously presented) A near-end speaker detection method according to claim 9, further comprising:

correlating the discrete signal spectrum of the delayed version of the digital audio signal with a visual speech activity estimate taken from a visual feature vector supplied by the visual feature extraction and analyzing device to provide a further estimate for updating the estimate for the frequency spectrum corresponding to the signal which represents said speaker's voice as well as a further estimate for updating the estimate for the noise power density spectrum of the statistically distributed background noise.

11. (Previously presented) A near-end speaker detection method according to claim 9, further comprising:

adjusting cut-off frequencies of a band-pass filter used for filtering the discrete signal spectrum of the digital audio sequence based on a bandwidth of the estimated frequency spectrum.

12. (Previously presented) A near-end speaker detection method according to claim 4, further comprising:

adding an audio speech activity estimate obtained by amplitude detection of a band-pass-filtered discrete signal spectrum of the digital audio sequence to a visual speech activity estimate taken from a visual feature vector supplied by said visual

feature extraction and analyzing device to provide an audio-visual speech activity estimate,

correlating the discrete signal spectrum with the audio-visual speech activity estimate to provide an estimate for a frequency spectrum corresponding to a signal which represents a voice of said speaker as well as an estimate for the noise power density spectrum of the statistically distributed background noise; and

adjusting cut-off frequencies of a band-pass filter used for filtering the discrete signal spectrum of the digital audio sequence based on a bandwidth of the estimated frequency spectrum.

13. (Currently amended) A telecommunication system, comprising:
a video-enabled phone;
a video-telephony based application running on the video-enabled phone; and
a video camera built-in to the video-enabled phone and pointing at a face of a speaker participating in a video telephony session,

wherein said video-telephony based application comprises:

an audio sequence detection device for detecting an analog audio sequence;

an audio feature extraction and analysis device for analyzing said analog audio sequence and extracting said audio features therefrom;

a video sequence detection device for detecting said video sequence;

a visual feature extraction and analysis device for analyzing the detected video sequence and extracting said visual features therefrom;

a noise reduction device for separating a speaker's voice from said background noise based on a combination of derived speech characteristics by removing said separated background noise from said analog audio sequence and outputting a speech activity indication signal comprising a combination of speech activity estimates supplied by said audio feature extraction and analysis device and said visual feature extraction and analysis device; and

a multi-channel acoustic echo cancellation device for performing a near-end speaker detection and double-talk detection algorithm based on the speech characteristics derived by said audio feature extraction and analysis device and said visual feature extraction and analysis device,

wherein said noise reduction device is further configured to subtract a discretized version of an estimated noise power density spectrum of said background noise from a discrete signal spectrum of an analog-to-digital converted version of said analog audio sequence, said estimated noise power density spectrum being based on said audio features, said visual features and said discrete signal spectrum.

14. (Previously presented) A telecommunication device equipped with an audio-visual user interface and including the noise reduction system according to claim 1.

15. (Previously presented) A telecommunication system configured to perform the near-end speaker detection method of claim 4.